DOCUMENT RESUME

ED 099 405 TM 004 058

TITLE	What Research Says to the Teacher: Evaluation and Reporting of Student Achievement.
INSTITUTION	National Education Association, Washington, D.C.
PUB DATE	74
NOTE	32p.
AVAILABLE FROM	NEA Publications, Order Department, The Academic Building, Saw Mill Road, West Haven, Connecticut 06516 (Stock No. 387-11870, $0.50)
EDRS PRICE	MF-$0.75 HC Not Available from EDRS. PLUS POSTAGE
DESCRIPTORS	*Academic Achievement; Behavioral Objectives; Criterion Referenced Tests; Educational Philosophy; *Educational Testing; Grades (Scholastic); *Grading; Self Evaluation; Standardized Tests; *Student Evaluation

ABSTRACT

The four parts of this report are: (1) the purposes of evaluation and reporting; (2) relationship to educational philosophy and methods of teaching, including historical development, present reporting practice and criticisms of these practices, and alternatives to conventional grading; (3) the best way to report student achievement; and (4) evaluation to improve instruction, with comments on standardized tests, criterion-referenced tests, self-evaluation, and promising trends in evaluation. (RC)

What Research Says to the Teacher

Evaluation and Reporting of Student Achievement

ED 099 405

PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL BY MICRO FICHE ONLY HAS BEEN GRANTED BY

NEA

TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER.

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

EM 001 158

nea
National Education Association
Washington, D.C.

NOTE

Previously published material used in this book may use the pronoun "he" to denote an abstract individual, e.g., "the student." We have not attempted to alter this material, although we currently use "she/he" in such instances.

—NEA Publishing

Copyright © 1974
National Education Association of the United States

NEA Stock No. 387-11870

Library of Congress Cataloging in Publication Data

National Education Association of the United States.
Evaluation and reporting of student achievement.

(What research says to the teacher)
1. Grading and marking (Students) I. Title.
II. Series.
LB3051.N3 1974 371.27 74-10908

CONTENTS

Introduction	4
Purposes of Evaluation and Reporting	5
Purposes of Marking	5
Purposes of Evaluation	6
Relationship to Educational Philosophy and Methods of Teaching	7
Historical Development	8
Present Reporting Practice	12
Criticisms of Present Reporting Practice	13
Alternatives to Conventional Grading	17
What's the Best Way To Report Student Achievement?	24
Evaluation To Improve Instruction	25
Standardized Tests	25
Criterion-Referenced Tests	27
Self-Evaluation	28
Promising Trends in Evaluation	29
Selected References	30

INTRODUCTION

Are parents, teachers, and students in your school satisfied with current methods of evaluating and reporting student achievement? Probably not. Surveys indicate that most school systems are currently engaged in, have recently completed, or are about to begin a review or alteration of their evaluation and reporting procedures. But changes are often short lived as one school or school system adopts a method another school has just abandoned. Many schools go around in circles, eventually returning to a reporting system that had been discarded sometime earlier.

Why, despite many years of experimentation and research do many problems and issues related to evaluation and reporting remain unsettled? How can the classroom teacher determine what evaluation and reporting system is best for her/his class? How can evaluation and reporting systems serve to improve teaching and facilitate learning?

To help find at least partial answers to these questions and to stimulate further study, this report is divided into four parts: (1) The purposes of evaluation and reporting, (2) Their development, past and present, in relation to different educational philosophies and methods of teaching, (3) The best way to report student achievement, and (4) The types of evaluation that can improve instruction.

PURPOSES OF EVALUATION AND REPORTING

One major reason for the great difficulty in finding an ideal evaluation and reporting system is the multiplicity of purposes and interests of different people in these functions. A comprehensive and widely quoted definition of purposes is William L. Wrinkle's, published in 1947 in his classic work, *Improving Marking and Reporting Practices*. His definition follows:

Purposes of Marking

> 1. *Administrative functions.* Marks indicate whether a student has passed or failed, whether he should be promoted or required to repeat the grade or course, and whether he should be graduated. They are used in transferring a student from one school to another and in judging candidates for admission to college. They may be used by employers in evaluating prospective employees.
>
> 2. *Guidance functions.* Marks are used in guidance and counseling in identifying areas of special ability and inability, in deciding on the advisability of enrolling the student in certain courses and keeping him out of others, and in determining the number of courses in which he may be enrolled.
>
> 3. *Information functions.* Marks are the chief means employed by the school in giving information to students and their parents regarding the student's achievement, progress, and success or failure in his school-work.
>
> 4. *Motivation and discipline functions.* Marks are used to stimulate students to make greater effort in their learning activities. They are used for the same purpose in determining eligibility to honors of many different kinds such as participation in school activities, eligibility to play on the team, membership in selected groups, the winning of scholarships, etc. (39, pp. 31-32)

These four classifications are not mutually exclusive; they overlap. For instance, the use of marks for awarding a scholarship provides motivation, but it also serves an administrative function. Thus, almost any one function may be under more than one classification.

Grading Is Not Evaluating

Much confused thinking has been caused by lack of distinction between evaluation and marks or grades. Very often in educational literature the terms have been used interchangeably.

Evaluation of student achievement assesses how well students have achieved curriculum objectives. Grading, or marking, is not evaluating. Grading is merely one way to *report* an evaluation. A smile, a frown, a spoken or written comment, for example, are other ways to report an evaluation, as is selecting a certain student to lead a classroom discussion. And a computer printout profile of a student's accomplishment of dozens of specific behavioral objectives is still another.

What's Good for the Goose . . .

An evaluative or reporting technique that suits one purpose or person may not suit others, or it may even be in direct conflict with other purposes or needs of other persons. For example, conventional letter grades, A-F, have served as an administrative convenience to colleges in selecting students. Studies have shown letter grades in high school to be a fairly reliable indicator of the kinds of grades students will get in college. (18) But the grades students get in school are not good predictors of job performance in the world of work outside the school. (15, 21)

Or, to take another example, motivation may be enhanced for students who receive high marks or grades but may be destroyed for students whose self-confidence is undermined by frequent low grades. (8)

Purposes of Evaluation

The distinction between evaluation and grading is emphasized in the 1967 yearbook of the Association for Supervision and Curriculum Development, *Evaluation as Feedback and Guide*. (4) This book is based upon the assumption that the major purpose of evaluation is to provide feedback upon which to base individual and institutional decisions to improve education. The authors state, "The test of an evaluation system is simply this: Does it deliver the feedback that is needed, when it is needed, to the persons or groups who need it?"

If any system of evaluation is to meet this test, it must—

1. Facilitate self-evaluation.
2. Encompass every objective valued by the school.
3. Facilitate learning and teaching.
4. Produce records appropriate to the purposes for which records are essential.
5. Provide continuing feedback into the larger questions of curriculum development and educational policy. (4, pp. 4-6)

The ASCD Yearbook Committee concluded that our present system of evaluation and reporting does not satisfy any of these criteria and should be replaced with an entirely new system of diagnostic teaching that provides for considerable student self-evaluation and cooperative evaluation by the school faculty.

RELATIONSHIP TO EDUCATIONAL PHILOSOPHY AND METHODS OF TEACHING

Different practices in evaluation and reporting result chiefly from different educational philosophies and goals and from different methods of teaching and learning.

If our educational goal is to convey a fixed body of knowledge to all students through teacher-directed methods, then a percentage mark on a written test indicating how much of the body of knowledge was acquired is appropriate. But if the goal of the school is to develop the full potential of each individual student largely through self-direction, other means of evaluation and reporting must be devised.

When percentage marks, national standardized tests, and letter grades became firmly entrenched in the first two decades of this century, the goals of our schools were largely subject oriented. High school and college programs were based on elitist, highly selective standards. It was assumed that the predominant, middle class culture should be the standard for all.

Since then American education has been washed by several different waves of philosophy and method, each creating pressures for different kinds of evaluation and reporting. Although these waves of change dominated certain periods of time, no one of them has ever covered all schools at the same time or to the same extent. Today, for example, some schools that are com-

pletely immersed in a wave of humanism may be seeking to eliminate grades or external evaluations altogether; some schools still concerned with implementing the new emphases on the subject matter disciplines are trying to find more precise ways of defining and evaluating prescribed learnings, while most schools are working under a combination of these and other approaches to education and are therefore seeking an appropriate combination of evaluation and reporting devices.

Historical Development

To place letter grades and other aspects of evaluation and reporting in perspective, it would be helpful to review how we got where we are now.

Marks and grades are relatively new in the history of education. In the early 1800's in the one-room school where children of all ages and abilities studied together, students often moved from one level of work to another in each subject as soon as the teacher determined that they had mastered the prerequisite knowledge or skill.

Some were slower than others, but this did not cause them to be labeled as failures. They just took longer or received more help, as required. The teachers' reports to students and parents consisted of spoken or written descriptions of what the students had learned. Students' knowledge was tested not for the purpose of assigning grades but to demonstrate progress and for those few who were planning to go to college to determine what additional instruction was necessary to prepare them for college work.

Competition for admission to college did not become a major problem until after the rapid expansion of public high school enrollments in the last quarter of the nineteenth century.

As elementary and high schools grew in size, the students were divided into different levels or grades in the effort to provide more efficient mass instruction. Percentage marks began to be widely used for two major purposes: to help teachers classify students and to help colleges screen their applicants.

The mark became the major basis for determining whether a student was to be promoted to the next level and whether she/he was to be admitted to college.

The first major challenge to the marking system came from a study conducted by Starch and Elliott in 1912 which demonstrated very dramatically the subjective nature of supposedly objective grades. It was found that when the same English test paper was submitted to different teachers for grading the numerical scores assigned varied as much as 47 points on a 100-point scale. (35) In a later study there was a range of 67 points in marks assigned by different teachers to the same geometry test! (34) As similar findings from both other studies and informal observation confirmed the lack of precision of percentage grades, educators began using scales that had fewer but larger categories. Some used a three-point scale: Excellent, Average, or Poor. Some used a five-point scale: Excellent, Good, Average, Poor, or Failing, symbolized by A, B, C, D, F. The five-letter grading system came to be the one most prevalent in high schools and infiltrated elementary schools as well.

Between 1900 and 1920 the emphasis was upon making the meaning of grades more objective and clear. During this period scientific development of objective tests began. The use of standardized tests based on national norms to screen men for military service during the First World War stimulated the use of such tests in schools.

The two main types of tests used were achievement tests in the major subjects and IQ tests. Many educators interpreted IQ scores to be permanent indicators of general intelligence, and IQ became synonymous with academic ability. In recent years, however, many studies have shown that IQ tests measure only a few aspects of intelligence, are subject to substantial change during a person's life as a result of her/his particular experiences, and are particularly invalid for minority groups whose cultural or language background is different from the "average American." Furthermore, IQ scores can sometimes be affected by merely studying for the test.

Paralleling the growth of standardized testing was a movement to make grading fairer by allocating grades for student achievement on the normal curve. According to one interpretation, 3 percent of students should receive A's, 24 percent B's, 46 percent C's, 24 percent D's, and 3 percent E's. Thus, try as they might, a certain percentage of students were doomed to fail. Furthermore, since the national distribution of abilities is not

evenly distributed in every classroom or school, students of like ability and achievement receive different grades according to where they may happen to live.

Some educators in the 1920's tried to eliminate rather than tinker with grades. Some schools substituted verbal descriptions of the students' abilities. Some tried pass/fail systems. Others advocated a mastery approach in which evaluation would seek to determine only whether or not a student has mastered an educational objective.

The Progressive Education movement beginning in the late 1920's and reaching its peak in the 1930's stimulated much questioning of and experimentation with various means of evaluation and reporting—seeking to foster growth of the total person, individual differences, and individualized instruction. The Progressives believed that education should bring about changes in behavior, not just facilitate the memorization of facts.

A comprehensive report on evaluation based on the Eight Year Study of the Progressive Education Association stated that (a) evaluation and recording must be directly related to each school's purposes and philosophy, (b) an evaluation program must be comprehensive, including appraisal of progress toward all the major objectives of the school, and (c) teachers should participate in the construction of all instruments for evaluation and forms for records and reports. (32)

The Progressive Education movement stimulated experimentation with anecdotal records, student-teacher and parent-teacher conferences, letters to parents, check forms, and evaluation in terms of achievement of educational objectives, including behavioral objectives.

During the 1930's, William L. Wrinkle and his colleagues at the Colorado State College at Greeley tried and discarded almost all of the alternate types of reporting discussed in this report. They concluded that evaluation of student achievement should be in terms of behavioral objectives. But they listed only 15 broad objectives on reports to parents, keeping the specific objectives in each subject for internal use. Interestingly, the reports included an evaluation of how well students had achieved the objectives for each subject on a five-point scale. Words were used, but their meaning could easily be translated into conventional A-F letter grades. The faculty had intended to keep these evaluations for

confidential student records only, but the students requested that they be sent to parents.

"You are fortunate that you can begin your work without having to go through ten years of learning the hard way," Wrinkle wrote to teachers interested in improving evaluation and reporting. "But don't get the idea that your job is going to be easy. It won't be. Often the understanding and application of a very simple idea is very difficult—because it is so simple!" (39, p. 92)

The Second World War interrupted the ferment of change and experimentation brewed by the Progressives. Science and math became weapons in the war and in the postwar struggles with the U.S.S.R. for power and prestige. There was a movement toward a return to fundamentals in education with emphasis on the subject rather than the learner. In many schools this was accompanied also by a return to the old, academically respectable letter grades as a presumed means of motivating students to work harder.

The major thrust of new development was to redesign the curriculums of the subject matter disciplines. Although professional educators were given some role in the revision process, control was in the hands of erudite scholars of the various disciplines. They sought to put more depth, more significant concept building, more systematic study of the structure of each discipline into the new curriculums.

Another significant postwar development was programmed instruction. Programmed instruction brought more than teaching machines to the schools. It brought the philosophical influence of behavioral psychologists, such as B. F. Skinner, who believe that a high degree of control can and should be imposed on individuals to shape their behavior to predetermined goals and that learning can be very systematic, sequential, and efficient. The behavioral psychologists gave new impetus to the application of behavioral objectives in curriculum planning and evaluation on a much more rigid basis than the objectives advocated by Tyler and other early Progressives.

Although the new curriculums and programmed systems of instruction have met with varying success, most of them failed to live up to the expectation that they would bring dramatic improvements to American education. Perhaps because of disap-

pointment, perhaps because they lost interest or hope, academic scholars allowed control of the schools to drift back into the hands of professional educators.

At any rate, during the 1960's, the influence of the academic elite was overwhelmed by the influences of the civil rights movement, which sought equal opportunity for minorities in education, by student activists seeking more relevant education, and by a renaissance of the basic tenets of Progressive Education under a new title: Humane Education.

Present Reporting Practice

In the 1970's, as more schools seek to reemphasize growth of the total person, respect for individual differences, and individualized instruction, they are attempting to devise and apply appropriate means of evaluating and reporting pupil progress. But as a glance at Table I will indicate, the same old means of reporting were generally in effect in 1971, when NEA Research last surveyed pupil progress reports to parents.

A classified scale of letters is the most popular procedure used in junior high (82.4 percent) and senior high schools (83.8 percent). Teacher-parent conferences appear to be most popular in kindergarten, used by 85.4 percent of the systems, and in grade 1, used by 77.2 percent of the systems. In grade 4, each procedure is used by about 70 percent of the systems. Note that some systems must use both procedures because of the relatively high percentages of each. Some of the other procedures must also be used in the same systems.

Descriptive word grades, although used by about one third of the systems in kindergarten and grade 1, are used infrequently in the junior and senior high schools. A formal letter or written paragraph to parents is more popular than descriptive word grades in the high schools. The other procedures are used by relatively small percentages of the systems at each level. The data showed no differences in the procedures used by different size systems.

Table I.—Pupil Progress Reports to Parents Used by School Systems for Selected Grade Levels, 1971.

Reporting procedure	Kinder-garten	Grade 1	Grade 4	Junior high	Senior high
Classified scale of letters (e.g. A-F)	14.0%	45.1%	71.6%	82.4%	85.8%
Teacher-parent conferences	85.4	77.2	69.6	48.0	39.7
Descriptive word grades (e.g., excellent, poor)	30.7	32.6	16.6	6.7	6.0
Formal letter or written paragraph to parents	27.5	23.0	17.8	14.3	16.2
Pass-fail	7.3	7.3	4.7	4.3	6.2
Classified scale of numbers (e.g., 1-5)	2.3	5.1	6.5	5.3	4.8
Percentage grades	0.8	2.8	4.4	8.7	11.1
Dual-marking system (e.g., A/S, 91/s, C/A)	0.8	4.0	5.1	6.1	4.1
Other	3.9	5.5	4.0	2.1	1.3
Estimated number of systems	8,722	11,017	10,983	10,492	10,194

Source:

National Education Association, Research. "Reporting Pupil Progress to Parents 1966 and 1971." *NEA Research Memo 1972-10.* Washington, D.C.: the Association, September 1972, p. 1.

Checklists, perhaps because they are generally considered to supplement a basic report, were not listed in this survey questionnaire, but other studies have shown that they are used by many schools at all levels.

Criticisms of Present Reporting Practice

Obviously there is wide diversity in reporting practices in schools of the United States. Such diversity is bound to cause problems in a society so mobile that one out of five families moves each year. But some educators take comfort from the situation by reasoning that the diversity indicates a willingness on the part of many schools to continue to experiment in a search for a procedure that describes adequately all the dimensions of a pupil's physical, mental, and emotional development. Furthermore, no other aspect of American schooling is uniform. As long

as we have great differences in educational philosophy and practice, we shall continue to have differences in evaluation and reporting of student progress.

The Tail That Wags the Dog

However, despite many variations, letter grades or marks definitely dominate our schools from elementary school through college. They dominate them not only as a means of evaluation and reporting, but in some schools they dominate the entire learning process. Many educators now believe that instead of being an adjunct or aid to the learning process, grades have become a major obstacle to teaching and learning. (17, p. 78)

What's Wrong with Letter Grades?

Following is a summary of five of the major negative criticisms of letter grades as a means of evaluation and reporting:

1. *Grades rather than learning have become the goal of students.* This has greatly narrowed the real curriculum of the school, encouraged cheating, and discouraged desirable learning attitudes and habits. (17)

2. *Grades do not tell us anything specific.* Since they are derived from a composite of many factors, we can't tell, for example, whether a student who receives a C in English can express thoughts very clearly but cannot spell correctly, whether the reverse is true, whether the child is strong in literature or weak in grammar, or whether she/he is average in all her/his English skills. (7)

3. *Grades have different meanings because of subjective factors.* For example, studies have shown that teachers generally tend to grade girls higher than boys and middle class children higher than lower class children for identical achievements. (10, 29)

4. *Grades have different meanings because they are based upon different criteria.* In different schools, and even within the same school or classroom, a grade may represent an attempt to measure achievement in relation to one's classmates, in relation to national norms, in relation to each child's ability, in relation to teacher expectations, in relation to pupil effort, or a combination of some or all of these factors. (39)

5. *Low grades have serious adverse effects on some students' self-concepts* that not only interfere with learning but have serious negative lifetime effects on personality and mental health. (8)

Recent Attacks on Marks and Grades

In recent years the grading system has been strongly attacked by organized groups as well as by individual researchers and writers. Here are just a few examples:

In the 1960's many student demonstrations and revolts on college campuses sought, among other things, elimination or reform of the grading system.

In 1967 the Yearbook Committee of the Association for Supervision and Curriculum Development, in *Evaluation as Feedback and Guide*, declared that the prevalent marking grading system is largely responsible for making our system of evaluation grossly inadequate in providing valid feedback to the individual learners themselves and to the larger units of the school system and for gross exaggeration of the more mechanical, easier-to-measure features of education. "The end result is not simply bad evaluation; it is distorted teaching and learning." (4, p. 15)

A staff report of the National Education Association, *Schools for the 70's and Beyond: A Call to Action*, published in 1971, declared, "Grading—the process of appraising youngsters in relation to one another, ranking them on a ladder from best to worst—is not only unnecessary to instruction but is often positively harmful." (24, p. 64)

And in the same year the Board of Directors of the National Council of Teachers of English declared:

> Reporting a child's progress in the early years should be done through methods other than the assignment of a letter or numerical grade. Rather, the reporting of a child's progress should be through regular conferences between teacher and student based upon anecdotal records, comparative samples of the child's own work, the teacher's estimates of the child's growth in skills, and his growth toward achieving other goals that the community and the school might have set. . . .

> After the early years, at all educational levels only passing grades . . . should be recorded on a student's permanent record. . . . (26)

Why Have Marks and Grades Persisted?

Why, despite these many criticisms, do letter grades persist as the major means of reporting student achievement?

1. *Inertia to change* anything in life makes people cling to familiar things even when they are not satisfied with them.

2. *Familiarity with grades* makes it difficult for parents, students, and teachers to conceive of evaluation in other terms, even when changes are attempted. Many educators who have tried to substitute another means of reporting to parents have had reactions like the following: After a 10-minute parent-teacher conference or after an examination of a descriptive letter or checklist of learnings achieved, a parent says, "This is very interesting, but what does it *really* mean? Is Maria a B student, a C student, or what?"

3. *Grades represent simple concepts*, and people usually prefer the simple to the complex.

4. *Grades are easy for teachers to prepare* compared to most other means of evaluation that have been tried. Several of the alternatives that have been proposed are so time-consuming that most teachers find it impossible to carry them out properly over an extended period of time.

5. *Grades are a spur to classroom performance* for students. Studies have shown that in conventional school situations when grades are eliminated for some students and retained for others, work output of the ungraded student decreases in comparison to the graded student's. (14)

6. *Grades are used to screen students for college admission.* Now that many parents want their children to go to college, they start worrying about whether their offspring will have the grades for college even before they enter school.

7. *No substitute for grades has won continuing support.* Except for a few limited applications, such as parent-teacher conferences in the kindergarten, substitutes for grades have proven to be too time-consuming, too difficult to use or understand, or have not found continuing acceptance for other reasons.

Alternatives to Conventional Grading

Following is a brief description and analysis of each of the major alternatives or supplementary means of reporting student achievement that schools are trying currently. Since most of them have been tried since the 1920's, considerable research and experience have accumulated to help us assess their advantages and disadvantages.

Dual Grades and Ratings

As indicated earlier, a single grade mark could mean any one of a number of things. It could indicate the achievement of a student in terms of an absolute scale; in terms of other students in her/his class, school, school system, or nation; or in terms of her/his own ability, growth, or effort. Some schools attempt to define clearly upon which of these factors a grade is based. Some schools deliberately combine several of the factors. And some teachers consciously or unconsciously combine factors even if it is not official policy.

To reduce such confusion, some schools grade on more than one basis. A common type of dual grading used in both elementary and high schools is the achievement/effort combination. Achievement is usually rated A-F and effort O for outstanding, S for satisfactory, or U for unsatisfactory.

Another type of dual grading, used mostly in elementary schools, employs one rating to indicate individual progress and another to indicate level of performance. For example, A/4 may indicate that a sixth-grade student has achieved outstanding growth in reading, but is reading at a fourth-grade level.

There are many other variations on this theme, with some schools giving three or more grades in each subject.

Theoretically, dual grading systems should be superior to single-grading systems because they describe more than one aspect of achievement. But studies have shown that there is a halo effect which causes teachers to tend to give students who achieve high or low grades in terms of grade norms similar grades in terms of ability. (13)

Rating Scales Other than Percentages or A-F

In efforts to reduce emphasis on working for grades and the discouraging effects of poor grades, some schools have used rating scales with only two or three steps and have tried to express poor ratings euphemistically.

Examples of such two-point rating scales are—

S—Satisfactory	or	G—Good Progress
U—Unsatisfactory		N—Needs To Improve

But there is a tendency to expand the number of ratings to three, four, or more, such as—

Outstanding
Satisfactory
Is Showing Growth
Needs Improvement.

This seems to be another case of history repeating itself, for as Wrinkle wrote in 1947:

> The experience of many schools that changed to the S and U marking system was that with the removal of the possibility of an A, B, or C many students became concerned only in staying just over the border in the S area. As a corrective move to recover the stimulation which the A had provided, they added a third letter, usually an H which stood for "Honors." Some affixed + and − signs to the H and S (H, H−, S+, S, S−), and they were back almost where they started; they then had a 6-point marking system.
>
> Progress in the improvement of marking and reporting practices cannot be achieved by the mere manipulation of symbols. About the best that can be said for the substituting of S and U or H, S, and U for A B C D F marks is that thereafter the school is brought face to face with the fact that what it thought was a problem in marking is fundamentally a curriculum problem. If students quit working when the incentive of marks is removed and the staff is unwilling to admit that they can be stimulated to learn only by the use of such extrinsic pressures, then the staff has discovered something fundamental. . . . (39, pp. 51-52)

Parent-Teacher Conferences

Many educators believe that parent-teacher conferences are the ideal method of reporting to parents. If properly handled, they can provide individualized, diagnostic, and tactful two-way communication. To reach this ideal the teacher needs to keep and refer to written evaluative records such as anecdotal records, tests, checklists of student achievements, and samples of pupils' work.

A written record of the conference should be kept for future reference. Some school systems provide for a written report to be prepared by the teacher in advance as a guide to discussion and for presentation to the parent during the conference.

Obviously, a major drawback of the parent-teacher conference is the great amount of time it takes. It is almost impossible to use it on a regular basis in secondary schools where each teacher may teach 100 to 150 students. Its use as the major means of reporting is generally limited to kindergarten and the first grade. However, it is being used increasingly at all levels to supplement written reports.

Narrative Reports

Narrative reports range from teacher-written letters to comments under various headings on a report card. Often narrative reports are alternated with parent-teacher conferences in kindergarten and primary grades. Such reports offer some of the same advantages and disadvantages as parent-teacher conferences. They can be tailored to the particular child and parent and can give meaningful details of progress and problems. However, writing a good letter or comment of this type is difficult, and teachers often resort to vague, general language.

Some schools have been experimenting with computers to help write letters to parents. The teacher, who knows the complex particularities of each student's achievement and potential, selects from a large number of precoded comments that have been programmed into the computer. The computer then prints out the letter or report form. When properly used, this method both saves a great deal of teacher time for instruction and permits detailed reporting of each student's progress. At present the major complaints against this method have been awkward combi-

nations of words and phrases, some gross errors, and under utilization of the sophisticated technology. If these flaws can be worked out, computers may become a generally useful and practical aid to effective recording and reporting of student achievement.

Checklists

Checklists are shortcuts to the writing of reports by teachers in which preprinted statements of student characteristics, learnings, or behavior are checked or marked with some other evaluative symbol. Using checklists is the simplest way to report the most information with the least expenditure of teacher time and effort. (39)

There are many types of checklists. They range from—

1. Vague descriptions of a few character traits and study habits supplementing conventional reports on academic subjects (_____ gets along well with others), through
2. Positive evaluations used to report what the student has achieved (_____ reads with understanding), to
3. Precise statements of behavioral objectives for all school subjects and goals: (Given a human skeleton, the student must be able to correctly identify by labeling at least 40 of the following bones: _____).

Behavioral Objectives

The use of behavioral objectives as a guide to teaching and evaluation has been one of the most publicized new concepts in education during the 1960's and early 1970's. Because of the difficulties many teachers have in implementing this much talked-about concept and because of several different interpretations and controversial issues, more space will be devoted here to analyzing this means of evaluation and reporting than to any of the other alternatives.

Purist proponents of behavioral objectives hold that all instruction can and should be planned to bring about controlled changes in behavior, that the learner must be able to demonstrate the change through performance, and that it is possible to observe and measure the success or efficiency of the student's performance.

Guides to writing behavioral objectives usually emphasize specificity, and most definitions of behavioral objectives follow Mager's three criteria.

A behavioral objective states—

- An action performed by a student.
- The condition under which the performance is to occur.
- The criteria of acceptable performance.

For example: "Given a human skeleton [condition], the student must be able to correctly identify by labeling [action] at least 40 of the following bones . . . [criteria]." (23)

According to Geis, "Most authors pay special attention to the *verb* in the statement of an objective. Such 'non-observable' verbs as think, appreciate, enjoy, know, are unacceptable while specific, observable action verbs (e.g., writes, assembles, states) are given a stamp of approval." (12)

However, many educational scholars such as Ralph Tyler, a pioneer in the study and use of behavioral objectives, include in the definition of behavior "all kinds of human reactions like thinking and feeling as well as overt reactions. . . . It includes attitudes toward subjects or things. It encompasses being able to solve problems and to acquire intellectual skills like reading or physical skills like running." (31)

It takes a large number of specific behavioral objectives to describe goals and accomplishments in any subject. For example, the goals for a semester of mathematics may take 100 or more objectives to describe. This specificity facilitates use of behavioral objectives for internal evaluation to improve learning. An objective stated specifically (e.g., "Given a set of figures, the student will identify those which contain a right angle.") is useful to the student in guiding her/his learning and developing self-evaluation, and is useful to the teacher in planning diagnostic teaching.

However, as Wrinkle found, it is probably inadvisable to report to parents in terms of all the detailed specific course objectives. Many of the objectives would not be understood by a number of parents, while those who did understand them might attempt inadvisable corrective measures. (39)

Some scholars believe that it is not desirable to use objectives that are as specific as those currently in vogue. Tyler says:

> I think many current uses of the term, *behavioral objectives*, imply procedures that are too specific. I believe that

> the individual human being is able to solve many of his own problems and so I think that more of our educational objectives should be general in nature—like learning how to go about attacking problems, finding out where the difficulties are, getting information, analyzing the data, and drawing inferences from data.
>
> Hence, in my view, many behavioral objectives should be set at a considerably higher or more general level than the extremely specific things I find in many current efforts to write them. (331, p. 42)

Other observers believe that it is not possible for teachers to use as many specific objectives as are generally advocated today. For example, Frances R. Link and Paul B. Diederich, in a proposal for a cooperative evaluation program, recommend reporting on schoolwide objectives plus the "four, five, or six most widely recognized objectives of each field of study—those that teachers are able and willing to measure in one way or another at this stage—not the long lists of 100 or more objectives that they often submit as claims that are never substantiated." (4, p. 129)

Several writers advocate different levels of specificity of objectives for different purposes. Three levels are most frequently suggested. Kratwohl, for example, classifies objectives as "global," "intermediate" (e.g., for a course), and "specific," (each representing a skill or concept) (20).

Even some of the most enthusiastic advocates of behavioral objectives have found that "there are some important goals which we have for our children which are *currently* unassessable. To the extent that such goals are extremely meritorious," states W. James Popham, "they are worth the risk of our pursuing them even if we cannot reliably discern whether they have been accomplished." Popham suggests that in some content-laden classes the proportion of nonmeasurable goals might be smaller than in other courses such as the humanities and aesthetics. (27, p. 608) Popham and others now tend to speak of instructional rather than behavioral objectives.

There are a number of educators who take a basic stand against the use of behavioral objectives for any education other than training to do a specific task or to acquire basic skills needed by all students. Some critics of behavioral objectives tend to believe that learning is or should be largely self-directed, unstruc-

tured, and in large part unpredictable. Some challenge the concept that all learning should or can lead *directly* or *immediately* to changes in behavior.

George F. Kneller says flatly, "Learning leads to no particular behavior. . . . To use behavioral objectives in individualized instruction is to overlook the essential differences between individual learning, knowing, and behaving." (19)

Robert L. Ebel writes that the real objectives of education are the knowledge and understanding, the attitudes and values that *induce* changed behavior now or in the future. Ebel concludes:

> Little that is wrong with any teacher's educational efforts today can be cured by getting him to define his objectives more fully and precisely. We ought not to ask teachers to spend much of their limited time in writing elaborate statements of their objectives. (9, p. 173)

Pass/Fail Or Credit/No Credit

The second most talked about current innovation in evaluation and reporting is pass/fail or credit/ no credit.

Pass/fail systems have been tried in colleges and high schools—

1. To encourage students to take difficult courses which they otherwise might avoid for fear of lowering their grade point average.

2. To allow students to apportion their study time according to their personal interests and needs, rather than spend most of their efforts in working for grades.

3. To foster free and creative approaches to learning.

4. To reduce the anxiety of poor achievers by encouraging them to concentrate on learning rather than on striving for a grade.

Of course the latter objective is not achieved when a student is in danger of failing. The credit/no credit system reduces such anxiety because no failing entry is made on the pupil's record.

Some credit/no credit plans also offer another advantage. They allow a flexible time period in which a student may complete the requirements of the course. Under this plan most students are assured not just of freedom from failure, but of eventual success.

Some people think of behavioral objectives and credit/no credit systems as opposites, representing control and freedom, respectively. But there is an increasing tendency toward marriage of these two seeming opposites. NEA's *Schools for the 70's* states that though "operationally stated objectives" and "evaluation on the basis of performance, not time," were developed independently, ". . . these ideas are not alternative ways of evaluating; rather in a complete scheme of evaluation, they can be used together within the instructional framework to diagnose—based on the concept that schools exist not to judge children, but to enable them to succeed." (23, p. 64)

One reason for resistance to pass/fail or credit/no credit systems has been the fear that colleges will not give favorable consideration to students who do not have conventional grade records. However, the authors of *WAD-JA-GET? The Grading Game in American Education* say that several surveys have indicated that most colleges are willing to accept teacher evaluation letters and student self-evaluations, in lieu of the usual grades and class rank. (17, p. 314)

The authors of *WAD-JA-GET?* present an interesting proposal. They suggest that a high school allow individual teachers and students to choose whether they wish to work in a conventional grading system or in credit/no credit courses. Teachers may be allowed to try some courses of each type if they wish, but students must choose all courses in either traditional grading or credit/no credit to avoid the possibility of taking unfair advantage of the situation by only taking difficult courses for credit/no credit. (17)

WHAT'S THE BEST WAY TO REPORT STUDENT ACHIEVEMENT?

The preceding discussion of the pros and cons of various means of reporting student achievement should make it clear that neither research, experience, nor theory have provided any single widely acceptable means of reporting. But though we find no simple solution, we do have some bases for making intelligent choices.

In most school situations a combination of several different types of reporting is needed to reflect current philosophy, prac-

tices, and expectations. Because they are so strongly entrenched in our schools and our culture, letter grades will probably continue to be used as one means of reporting pupil progress even in situations where other means of reporting that are more consistent with current philosophy, goals, and teaching methods are used also.

Classroom teachers have often felt frustrated because they had no choice or voice in selecting means of reporting, but increasingly, teachers are represented in committees considering schools- or system-wide changes in reporting and sometimes they are even given an individual choice. Even if you have no choice, you can make the best of the reporting system you have by understanding its meaning, strengths, limitations and relationships to your teaching program and by helping students and parents to share such understanding. You can try also to use means of evaluation in your classroom not only to provide data for reporting but also as a means to improve instruction.

EVALUATION TO IMPROVE INSTRUCTION

Evaluation and reporting are so closely interrelated that it is difficult to consider them entirely separately. Thus, several major issues regarding evaluation have been discussed in the preceding sections on alternative reporting practices, particularly in the discussion of behavioral objectives. But, as emphasized early in this report, evaluation should be considered primarily as a means of constant feedback for improvement of instruction rather than merely as a basis for occasional reports of student progress.

Although teachers use observation of class participation, anecdotal records, and student accomplishments in projects and reports for evaluation, tests—informal or teacher-made and standardized—have usually served as the most frequently used means of evaluation.

Standardized Tests

Until recently national standardized tests dominated the evaluation scene and greatly influenced the nature of teacher-made tests. Intelligence, aptitude, and achievement tests were all normative-based. These tests were designed to determine how

students ranked in relation to the national average with an equal number of scores distributed, according to the normal curve, above and below the median. Thus, in normative-based testing, half of the students must be below normal. If teaching and learning were to improve so that most children reached or surpassed the median score on a particular test, the test would have to be made harder so that half the students taking the test would always fail. (6)

During the 1960's and early 70's IQ tests were severely criticized because they are "biased against those who are economically disadvantaged and culturally and linguistically different, and especially against all minority groups." (26, p. 26) Furthermore, ". . . the use of the typical intelligence test contributes to what has come to be termed 'the self-fulfilling prophecy,' whereby students' achievement tends to fulfill the expectations held by others." (37, p. 54)

As increased emphasis has been placed on writing specific learning objectives for local school systems, schools, classrooms, and even individual students, standardized tests have become ever more unsatisfactory as evaluative instruments, for they tell more about how students rank in relation to each other than what they have accomplished in relation to their goals.

Dissatisfaction with standardized tests became so strong by 1972 that a resolution was passed that year by the NEA Representative Assembly to encourage "the elimination of group standardized intelligence, aptitude, and achievement tests to assess student potential or achievement until completion of a critical appraisal, review, and revision of current testing programs." A Task Force on Testing was appointed to study the situation and to prepare a final report for the 1975 Assembly.

An interim report of the Task Force in 1973 concluded with this statement:

> In summary, the Task Force believes that the major use of tests should be for the improvement of instruction–for diagnosis of learning difficulties and for prescribing learning activities in response to learning needs. They must not be used in any way that will lead to labeling and classifying of students, for tracking into homogeneous groups as the major determinants to educational programs, to perpetuate an elitism, or to maintain some groups and indi-

viduals "in their place" near the bottom of the socioeconomic ladder. In short, tests must not be used in ways that will deny any student full access to equal educational opportunity. (37, p. 54)

Criterion-Referenced Tests

Partially in response to such criticisms and partially in response to new developments in curriculum and methods of instruction, such as individualized instruction, national test makers as well as state and local school systems are beginning to develop new kinds of tests called "criterion-referenced tests." Criterion-referenced tests, or the similar "objectives-referenced" tests, have also been advocated and used lately for evaluating the effectiveness of teaching as part of the accountability movement.

Criterion-referenced tests seek to determine how well students have mastered their specific learning objectives rather than how they rank with other students. Their major purpose is to measure individual progress and identify needed additional experiences to improve learning.

The National Assessment of Educational Progress utilizes criterion-referenced tests. The individual teacher who uses very specific learning objectives as a basis for creating teacher-made tests is also utilizing criterion-referenced testing.

William F. Brazziel describes the potential for use of criterion-referenced tests as follows:

> In an improved situation utilizing CRT, the teacher would administer a locator or formative test using material to be mastered as criteria and would teach the children for a year. Over this period, he would have administered many summative tests indicating mastery of the units and skills he set out to teach during the year. He might give a final comprehensive summative test. Under this system, the children would end the year with comprehensive, explicit records that would indicate what they set out to learn, what they had learned, and the point at which learning should begin in the next school year (6, p. 53)

However, in actual practice the distinction between criterion-referenced and norm-referenced tests may be quite blurred. Some test makers use very similar procedures to construct items, often

the same items, and use test statistics designed for norm-referenced items in order to select items for criterion-referenced tests. There are no clearly defined and commonly agreed upon procedures for constructing criterion-referenced tests, and in fact many criterion-referenced tests are actually norm-referenced tests in disguise. The distinction is often a matter of emphasis rather than being clear-cut.

Some writers claim that neither criterion-referenced tests nor objectives-referenced tests eliminate the most common deficiencies of tests in general. They both still measure short-term retention and simple tasks at the expense of long-term retention, relearning abilities, and higher-level thought processes. (33) Complex performances are so difficult to measure that simpler tasks are substituted as test items. Binet's own categories of mental imagery, imagination, aesthetic appreciation, and moral sensibility are left unmeasured.

Self-Evaluation

Even more significant than new kinds of tests are recent trends in school programs that depend less upon formal testing of any kind and more upon student self-evaluation as an integral part of the learning process. Individualization, flexible scheduling, open schools, humane education, diagnostic teaching, and other programs that seek to tailor the learning situation to the specific needs and styles of specific learners make possible, in fact require, a high degree of student self-evaluation.

Several conditions are necessary for student self-evaluation to be effective. The student should be in a climate that encourages self-direction. Each student should have a role in establishing her/his learning goals, and the teacher should furnish instructional materials that have built-in evaluation aids, e.g., balance beams, new kinds of programmed books, and old-fashioned answer books.

Another important condition for effective self-evaluation is the opportunity for frequent individual student-teacher conferences, for self-evaluation does not eliminate the need for teacher guidance or teacher evaluation. In fact, teachers who promote student self-direction and self-instruction are cautioned to prepare students for such responsibility gradually and systematically.

Doris M. Lee supports student self-direction as follows:

> As the children or adolescents in a group grow and learn, it is totally impossible for any teacher to know just what is most needed by each and every one. For years we have assumed that they all need about the same thing. . . . We now know that the necessary alternative is to come closer to helping each move ahead fairly continuously at his own growing edge. Yet we are still faced by the virtually insoluble problems of providing so much differentiation if the teacher has to plan it all. Self-direction, within a reasonable framework, seems the most effective solution. For it adds the learner's perceptions and diagnosis of his own case to the teacher's broader understandings of the field as a whole. (22, p. 77)

Lee says further, "If, when they are on their own, tomorrow's citizens are to keep themselves knowledgeable and informed so that they may continue to be effective citizens of their world, they must begin to learn self-direction now." (22, p. 76)

Promising Trends in Evaluation

Ole Sand summarizes promising trends in evaluation as being:

From	To
1. tests as punishment	evaluation as a stimulant, a humane guide to continued growth and learning
2. measurement by paper and pencil tests	a variety of evaluation techniques with emphasis on observation
3. memory of the facts	focus on creativity and inquiry
4. exams at the end of a course	cooperative and continuous evaluation
5. narrow range of behaviors measured	evaluation of cognitive, affective, and psychomotor behaviors
6. evaluation only by the teacher	self-evaluation
7. colleges setting "standards" for admission	colleges cleaning up their sterile programs and working with schools to develop valid evaluation techniques. (30, p. 123)

SELECTED REFERENCES

1. Ammons, Margaret. "Objectives and Outcomes." *Encyclopedia of Educational Research.* Fourth edition. New York: Macmillan Co., 1969. pp. 908-14.
2. Anderson, Robert, and Ritsher, Cynthia. "Pupil Progress." *Encyclopedia of Educational Research.* Fourth edition (Edited by Robert L. Ebel.) New York: Macmillan Co., 1969. pp. 1050-60.
3. Armstrong, Robert J., and others. *Development and Evaluation of Behavioral Objectives.* Worthington, Ohio: Charles A. Jones Publishing Co., 1970. 99 pp.
4. Association for Supervision and Curriculum Development. *Evaluation as Feedback and Guide.* 1967 Yearbook. (Edited by Fred T. Wilhelms.) Washington, D.C.: the Association, 1967. 282 pp.
5. Baker, Eva L. "Instructional Objectives Exchange: Assistance in Goal-Referenced Evaluation." *Journal of Secondary Education* 45: 158-62; April 1970.
6. Brazziel, William F. "Criterion-Referenced Tests—Some Trends and Prospects." *Today's Education* 61: 52-53; November 1972.
7. Chansky, Norman M. "A Note on the Grade Point Average in Research." *Educational and Psychological Measurement* 24: 95-99; January 1964.
8. Child, I. L., and Whiting, John W. M. "Determinants of Level of Aspiration: Evidence from Everyday Life." *Journal of Abnormal Psychology* 44: 303-14; July 1949.
9. Ebel, Robert L. "Behavioral Objectives: A Close Look." *Phi Delta Kappan* 52: 172; November 1970.
10. Edmiston, R. W. "Do Teachers Show Partiality Toward Boys or Girls?" *Peabody Journal of Education* 20: 234-38; January 1943.
11. Gagné, Robert M. "Behavioral Objectives? Yes!" *Educational Leadership* 29: 394-96; February 1972.
12. Geis, George L. "Behavioral Objectives: A Selected Bibliography and Review." *SRIS Quarterly* 5: 19-20; Fall 1972.
13. Halliwell, Joseph W., and Robitaille, Joseph P. "The Relationship Between Theory and Practices in a Dual Reporting Program." *Journal of Educational Research* 57: 137-41; November 1963.

14. Hawk, T. L., and DeRidder, L. M. "A Comparison of the Performance of Pre-Graded Students with Grade-Motivated Students." *Journal of Educational Research* 56: 548-50; July-Augu 1963.
15. Hills, J. R.; Klock, J. A.; and Bush, M. "The Use of Academic Prediction Equations with Subsequent Classes." *American Educational Research Journal* 2: 203-206; November 1965.
16. Hoyt, P. Donald. *The Relationship Between College Grades and Adult Achievement.* ACT Research Report No. 7. Iowa City, Iowa: American College Testing Program, 1965. 57 pp.
17. Kirschenbaum, Howard; Simon, Sidney B.; and Napier, Rodney W. *WAD-JA-GET? The Grading Game in American Education.* New York: Hart Publishing Co., 1971. 320 pp.
18. Klugh, Henry E., and Bierley, Robert. "The School and College Ability Test and High School Grades as Predictors of Achievement." *Educational and Psychological Measurement* 19: 625-26; Winter 1959.
19. Kneller, George F. "Behavioral Objectives? No!" *Educational Leadership* 29: 397-400; February 1972.
20. Krathwohl, David R., and others. "Stating Objectives Appropriately for Program, for Curriculum, and for Instructional Material Development." *Journal of Teacher Education* 16: 83-92; March 1965.
21. Lavin, D. E. *The Prediction of Academic Performance.* New York: The Russell Sage Foundation, 1965. 182 pp.
22. Lee, Doris M. "Teaching and Evaluation." *Evaluation as Feedback and Guide.* 1967 Yearbook. (Edited by Fred T. Wilhelms.) Washington, D.C.: Association for Supervision and Curriculum Development, 1967. Chapter 4, pp. 72-100.
23. Mager, Robert F. *Preparing Instructional Objectives.* Belmont, Calif.: Fearon Publishers, 1962. 60 pp.
24. National Education Association, Center for the Study of Instruction. *Schools for the 70's and Beyond: A Call to Action.* Staff report. Washington, D.C.: the Association, 1971. 148 pp.
25. National Education Association, Research Division. *Marking and Reporting Pupil Progress.* Research Summary 1970-S1. Washington, D.C.: the Association, 1970. 55 pp.
26. National School Public Relations Association. *Grading and Reporting: Current Trends in School Policies and Programs.* Arlington, Va.: the Association, 1972. 64 pp.

27. Popham, W. James. "Must All Objectives Be Behavioral?" *Educational Leadership* 29: 605-608; April 1972.
28. Popham, W. James, and Baker, Eva L. *Establishing Instructional Goals*. Englewood Cliffs, N.J.: Prentice-Hall, 1970. 130 pp.
29. Rosenthal, Robert, and Jacobsen, Lenore. *Pygmalion in the Classroom: Teacher Expectation and Pupils' Intellectual Development*. New York: Holt, Rinehart & Winston, 1968. 240 pp.
30. Sand, Ole. "What To Teach—and How." *High School 1980: The Shape of the Future in American Secondary Education*. (Edited by Alvin C. Eurich and the staff of the Academy for Educational Development.) New York: Pitman Publishing Corp., 1970. pp. 122-23.
31. Shane, June G., and Shane, Harold G. "Ralph Tyler Discusses Behavioral Objectives." *Today's Education* 62: 41-46; September-October 1973.
32. Smith, Eugene R., Tyler, Ralph W., and the Evaluation staff. *Adventure in American Education: Appraising and Recording Student Progress*. New York: Harper and Brothers, 1942. Vol. 3, 550 pp.
33. Stake, Robert E. "Measuring What Learners Learn." *School Evaluation: the Politics & Process*. (Edited by Ernest R. House.) Berkeley, California: McCutchan Publishing Corp., 1973. 331 pp.
34. Starch, Daniel, and Elliott, Edward C. "Reliability of Grading Work in Mathematics." *School Review* 21: 254-95; April 1913.
35. ——. "Reliability of the Grading of High School Work in English." *School Review* 20: 442-57; September 1912.
36. Thomas, R. Murray. "Records and Reports." *Encyclopedia of Educational Research*. Fourth edition. (Edited by Robert L. Ebel.) New York: Macmillan Co., 1969. pp. 1104-10.
37. Today's Education. "NEA Task Force on Testing." *Today's Education* 62: 54; January-February 1974.
38. Vargas, Julie S. *Writing Worthwhile Behavioral Objectives*. New York: Harper & Row, 1972. 175 pp.
39. Wrinkle, William L. *Improving Marking and Reporting Practices in Elemeniary and Secondary Schools*. New York: Holt, Rinehart & Winston, 1947. 120 pp.